



Article

A Model Based on Fractional Brownian Motion for Temperature Fluctuation in the Campi Flegrei Caldera

Antonio Di Crescenzo ^{*,†} , Barbara Martinucci [†] and Verdiana Mustaro [†]

Dipartimento di Matematica, Università degli Studi di Salerno, 84084 Fisciano, Italy;
bmartinucci@unisa.it (B.M.); vmustaro@unisa.it (V.M.)

* Correspondence: adicrescenzo@unisa.it

† These authors contributed equally to this work.

Abstract: The aim of this research is to identify an efficient model to describe the fluctuations around the trend of the soil temperatures monitored in the volcanic caldera of the Campi Flegrei area in Naples (Italy). This study focuses on the data concerning the temperatures in the mentioned area through a seven-year period. The research is initially finalized to identify the deterministic component of the model given by the seasonal trend of the temperatures, which is obtained through an adapted regression method on the time series. Subsequently, the stochastic component from the time series is tested to represent a fractional Brownian motion (fBm). An estimation based on the periodogram of the data is used to establish that the data series follows an fBm motion rather than fractional Gaussian noise. An estimation of the Hurst exponent H of the process is also obtained. Finally, an inference test based on the detrended moving average of the data is adopted in order to assess the hypothesis that the time series follows a suitably estimated fBm.

Keywords: fractional Brownian motion; stochastic model; regression; time series; residuals analysis; Hurst exponent; Campi Flegrei caldera; temperature fluctuation



Citation: Di Crescenzo, A.; Martinucci, B.; Mustaro, V. A Model Based on Fractional Brownian Motion for Temperature Fluctuation in the Campi Flegrei Caldera. *Fractal Fract.* **2022**, *6*, 421. <https://doi.org/10.3390/fractalfract6080421>

Academic Editor: Mario Abundo

Received: 4 July 2022

Accepted: 27 July 2022

Published: 30 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In several applied contexts, one of the main problems in the description of a phenomenon that evolves over time is the individuation of the probabilistic laws by which the phenomenon itself is driven. To this aim, schemes based on the superposition of deterministic and random components are often chosen, in which the deterministic part describes the phenomenon's trend, while the random one describes the fluctuations determined by unpredictable exogenous factors. An example of such a model is presented in the work by Sebastiani and Malagnini [1], where a physical model with non-constant variance is proposed to describe the phenomenon of coupling erosion that caused the 2011 earthquake in Tohoku-Okai (Japan). Analogously, another compound model can be found in the work of Giordano and Morale [2]. Here, the authors propose a mathematical model for the prices of electricity in the Italian market through the years, whose stochastic component is given by the sum of a fractional Ornstein–Uhlenbeck process and the solution of a mean-reverting SDE driven by a Hawkes-marked process.

The aim of the present study is to implement an analogous scheme to describe the soil temperatures observed on the surface of the Campi Flegrei caldera. This volcanic area is located near the city of Naples in Italy and is famous for the ground deformations that have been registered in the area since ancient times. Recent observations, carried out from 2011 to 2017, were led to monitor the presence of radon in two sites of the area (cf. Sabbarese et al. [3]). In [3], the presence of radon in the soil was tracked, along with its dependence on physical quantities such as temperature, pressure and humidity, in order to prove how the seismic activity in the Campi Flegrei area was influenced by the presence of the element.

Starting from the research mentioned above, in this work, we aim to build a model able to describe the fluctuations in soil temperatures in the caldera through the identification of the temporal trend and the corresponding random component.

What emerged from the present study is that the deterministic component is described well by a piecewise linear model, where the segments' slopes are alternating in sign. The slopes can be determined by applying the ordinary least squares (OLS) method on the observed data. The endpoints for each segment of the piecewise curve are chosen through an empirical method. The stochastic model describing the random fluctuations component is instead recognized to be the fractional Brownian motion (fBm). For a description of the most relevant probabilistic and statistical aspects of fBm, see, for example, the works of Nourdin [4] and Prakasa-Rao [5]. This stochastic process has been proposed in several studies for modeling various geophysical phenomena. In the work by Mattia et al. [6], fBm is proposed as an approximation of the process describing the trend of ground data involving soil roughness collected over three different European sites. In the paper by Yin and Ranalli [7], the authors show that the event of earthquake rupturing in a fault is related to factors such as static shear stress and static frictional strength through a potential dynamic stress drop. The latter is regarded as a one-dimensional stochastic process, and it is modeled as an fBm with a Hurst exponent close to zero.

The evidence of the reasonableness of the model first emerges as consequence of a statistical test, in which the hypothesis that the random fluctuations of the model can be described by Brownian motion is rejected. The subsequent phase of the work involves the study of the periodogram of the data. It allows verifying that the random fluctuations are described well by fBm, in opposition to a fractional Gaussian noise (fGn). Such analysis also leads to the estimation of the involved parameters, namely the Hurst exponent H and the scale parameter D of the fBm. Many estimation techniques for the former parameters have been proposed in the literature, such as the methods presented in [8]. Other works often provide modified versions of the already known algorithms.

Subsequently, this study moves toward analyzing the residual series emerging from the considered stochastic process and simulated fBm with the previously estimated parameters. First of all, such a residual series is plotted with respect to the temperatures series, which suggests the Gaussianity of the residuals. Then, the procedure is repeated through suitable simulations. Two statistical tests are performed on the obtained values, with both leading to the acceptance of the null hypothesis of Gaussianity.

The goodness of the considered model is finally confirmed through an inference test, in which the test statistics involve the detrending moving averages obtained from the data.

The main results of this article are as follows:

- The development of a suitable stochastic model $X(t)$ describing a geophysical phenomenon;
- The estimation of the parameters involved in the model based on real observed data;
- The design of revised statistical algorithms to construct the model and to perform appropriate testing for confirming its validity.

Here is the plan of the article. In Section 2, we introduce the starting equation for modeling the soil temperature trend (cf. Equation (1)). Section 3 focuses on the data analysis directed toward identifying the deterministic component of the temperature trend. The remaining part of the article is dedicated to the analysis of the stochastic component of the model. More specifically, Section 4 is aimed at studying the periodogram obtained from the data observations, from which emerges the eligibility of fBm to describe the random component of the temperature compared to fGn. An estimation of the relevant parameters of the process (i.e., the Hurst exponent and the scale parameter) is also performed. Testing on the residuals of the model is then performed by resorting to the Shapiro–Wilk test and the robust Jarque–Bera test. The work is completed by Section 5, which contains the statistical test leading to the acceptance of the model, and Section 6, which collects the conclusions of the study.

2. The Stochastic Model

The time series considered throughout this study describes the temperatures of the surface soil of the Campi Flegrei caldera, obtained via daily observations in the time period between 1 July 2011 and 31 December 2017. The main aim of the research is to develop a model able to describe the fluctuations of the data. As usually observed in geophysical studies, the trajectory of the time series is seen as the superposition of a deterministic component and a stochastic one, where the latter represents the fluctuations from the trend. As a consequence, the process $\{X(t), t \geq 0\}$ that represents the daily observed soil temperatures is described by the following equation:

$$X(t) = r(t) + B_H(t), \quad (1)$$

where $r(t)$ constitutes the deterministic component of the process, while $B_H(t)$ is a stochastic process. We will show that a suitable assumption is to identify $B_H(t)$ as fBm. A similar model has been proposed for the vertical motion of the soil in the Campi Flegrei area (cf. Travaglino et al. [9]).

Under the assumption that $B_H(t)$ is indeed fBm, it is easy to observe that the model shown in Equation (1) is such that $\mathbb{E}(X(t)) = r(t)$. We recall, in fact, that fractional Brownian motion is a continuous-time Gaussian process with zero mean and covariance:

$$\text{Cov}[B_H(t), B_H(s)] = D(t^{2H} + s^{2H} - |t - s|^{2H}), \quad t, s \geq 0, \quad (2)$$

where $D > 0$ is a scale parameter named the diffusion constant and $0 < H < 1$ is a parameter known as the Hurst exponent. The process was first introduced by Mandelbrot and Van Ness in [10]. It is a generalization of Brownian motion, since for $D = 1/2$ and $H = 1/2$ it reduces to a standard Wiener process. We remark that fBm has the property of self-similarity; that is, for any choice of $a > 0$, the following is true:

$$\{B_H(at), t \geq 0\} \stackrel{d}{=} \{a^H B_H(t), t \geq 0\},$$

meaning that the two processes are equal in distribution. The parameter H is therefore also referred to as the scaling exponent or fractal index of the process.

The model in Equation (1) is not stationary, since fBm is not a stationary process itself. The process of the increments of $B_H(t)$, defined by $Z(t) = B_H(t+1) - B_H(t)$ for all $t \geq 0$, is, however, stationary. This is known as fractionary Gaussian noise (fGn).

Introducing the process $Z(t)$ leads to another well-known property of fBm, which is long-range dependency (LRD). As pointed out in [11], a stationary process $X(t)$ has long-range dependence (or is a long memory process) if its autocorrelation function, defined as

$$\rho(k) = \frac{\text{Cov}(X(t), X(t+k))}{\text{Var}[X(1)]},$$

satisfies the condition

$$\sum_{k=-\infty}^{+\infty} \rho(k) = \infty.$$

This definition implies that the autocorrelation of the process decays slowly over time, making the sum of the autocorrelations divergent. Therefore, if $X(t)$ is an LRD process, then

$$\lim_{k \rightarrow +\infty} \frac{\rho(k)}{ck^{-\alpha}} = 1, \quad (3)$$

where $c > 0$ and $0 < \alpha < 1$ are constants. This definition states that the decay of the autocorrelation function is power-like and hence slower than exponential decay. In the case of fBm, long-range dependence can be seen when looking at the increments in $Z(t)$. Moreover, the parameter α is related to the Hurst exponent through the equation $\alpha = 2H - 2$, evidencing

that the value of the Hurst exponent can be used to determine the nature of the process. We recall the following (see also [11,12]):

- If $\frac{1}{2} < H < 1$, then the increments of the process are positively correlated, making the process persistent (i.e., likely to keep the trend exhibited in the previous observations);
- If $0 < H < \frac{1}{2}$, the increments of the process are negatively correlated, and the process is counter-persistent (i.e., likely to break the trend followed in the past).

A relation similar to Equation (3) but concerning the frequency domain of the time series is shown in Section 3, where it is used for the estimation of a parameter of the process. It is worth mentioning that recent investigations showed that fBm with a random Hurst exponent can be employed to describe biological phenomena such as particle motion and intracellular transport, as reported by Balcerek et al. [13] and Han et al. [14], respectively.

3. The Deterministic Component

The model in Equation (1) was adopted to analyze the time series of the recorded temperatures of the Campi Flegrei caldera soil (cf. Sabbarese et al. [3]). The initial dataset consisted of $\hat{N} + 1 = 2005$ observations collected at the times indicated by $t_0, t_1, \dots, t_{\hat{N}}$ and shown in the scatter plot in Figure 1.

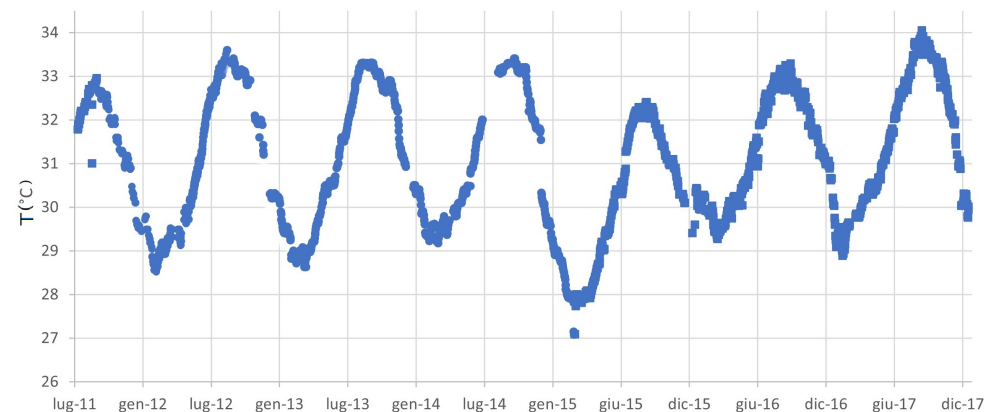


Figure 1. Scatter plot of the data series $X(t)$ for the temperatures of the surface soil of the Campi Flegrei caldera at the Monte Olevano (NA) site, collected during the observation period.

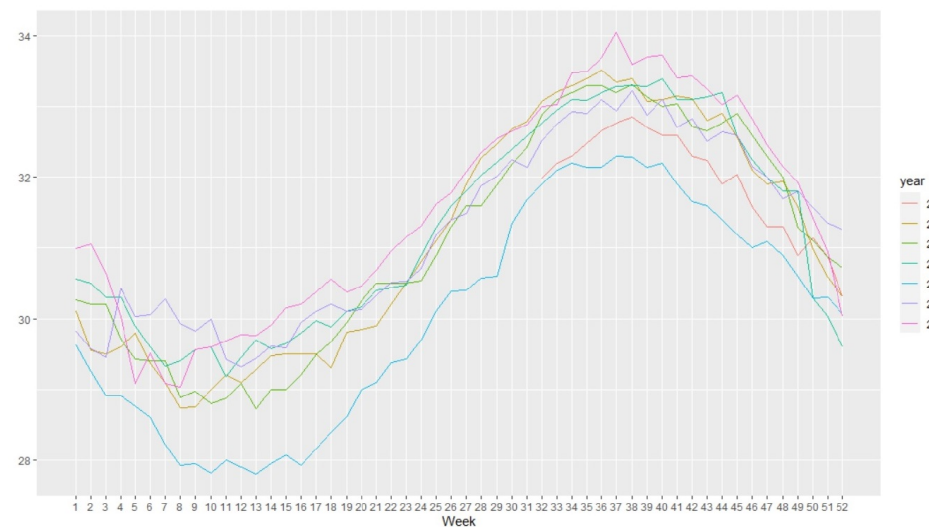
The observations are not equally spaced time-wise, since the data series was collected daily with the lack of a quite large number of observations. As shown in the plot in Figure 2, a seasonal trend is clearly visible in the dataset, with the temperatures increasing during spring and summer in each year and then starting to decrease around September. For this reason, the deterministic trend $r(t)$ is constructed by alternating segments with opposite slopes, which were obtained from the data through the OLS method. The first and the last measurement of the dataset, which are also the extremes of the curve, are given by

$$\begin{aligned} t_0 &= 07/29/2011, & X(t_0) &= 31.78; \\ t_{\hat{N}} &= 12/31/2017, & X(t_{\hat{N}}) &= 29.94. \end{aligned}$$

To obtain the points in which the curve changes its slope, a heuristic method based on linear regression was used, which was developed upon the methodologies illustrated by Hudson in [15]. First, for the dataset, we detected the local minima m_i , $i = 1, \dots, 6$ and maxima M_i , $i = 1, \dots, 7$, as shown in Table 1.

Table 1. Local maxima M_i and minima m_i for each subset of the dataset and the corresponding dates τ_i (day, month, and year).

τ_i	$M_i = X(\tau_i)$ (°C)	τ_i	$m_i = X(\tau_i)$ (°C)
16 September 2011	31.78	14 February 2012	28.68
25 August 2012	33.6	22 March 2013	28.63
11 September 2013	33.31	4 March 2014	29.17
22 September 2014	33.4	27 February 2015	27.08
5 September 2015	32.4	11 March 2016	29.27
19 September 2016	33.29	03 February 2017	28.88
30 August 2017	34.04		

**Figure 2.** Seasonal plot of the temperatures described by $X(t)$. This was obtained by taking weekly spaced values from the time series $X(t)$ and representing each year of observation (2011–2017) separately.

The time t_0 was used as the first extreme of the first segment of the piecewise linear curve. Subsequently, the slope α_j and the intercept of the regression line between τ_0 and the first local maxima M_1 were estimated using the OLS method. The corresponding coefficient of determination R^2 was also calculated. The second step consisted of repeating the operation and choosing the second extreme among the five values that preceded and the five ones that followed M_1 . Hence, among the 11 possible choices, we chose the regression line that maximized R^2 . Table 2 shows the values obtained for each endpoint.

Table 2. Slopes of the segments obtained by linear regression between initial time t_0 and a neighborhood of M_1 with their corresponding coefficients of determination. The slope was chosen such that R^2 was the highest, as underlined.

t_j	α_j	R^2
11 September 2011	0.0219	0.9302
12 September 2011	0.0217	0.9326
13 September 2011	0.0215	0.9341
14 September 2011	0.0213	0.9369
15 September 2011	0.0213	0.9405
16 September 2011	<u>0.0213</u>	<u>0.9442</u>
23 September 2011	0.0199	0.9021
24 September 2011	0.0190	0.8787
25 September 2011	0.0182	0.8574
26 September 2011	0.0174	0.8369
27 September 2011	0.0162	0.7880

Then, the same method was applied for the next 12 segments, in which the first initial time was determined by the previous step. The last regression line was finally calculated using $t_{\hat{N}}$ as the endpoint. We denote by $c_i > 0$ and $v_i < 0$ the alternating slopes obtained through the regression model. The endpoints θ_i for each interval, as well as the estimated slopes and the coefficients of determination, are reported in Table 3.

The values obtained by means of this procedure allowed us to identify the deterministic component $r(t)$ of the model in Equation (1). The values of R^2 confirm that the regression lines provided a good fitting of the data in each interval. The scatter plot and the resulting linear model are shown in Figure 3.

Table 3. Estimated slopes c_i and v_i and corresponding coefficients of determination evaluated for each subset $[\theta_{i-1}, \theta_i]$ of the data series, with $\theta_0 = t_0 = 29$ July 2011.

θ_i	c_i	R^2	θ_i	v_i	R^2
16 September 2011	0.0213	0.9442	27 February 2012	−0.0264	0.9709
13 September 2012	0.0262	0.9356	7 March 2013	−0.0311	0.9722
6 September 2013	0.0287	0.9818	7 March 2014	−0.0256	0.9573
24 September 2014	0.0237	0.9701	1 March 2015	−0.0418	0.9635
23 August 2015	0.0294	0.9659	19 March 2016	−0.0145	0.9287
8 September 2016	0.0238	0.9619	16 February 2017	−0.0261	0.9357
14 September 2017	0.0228	0.9674	31 December 2017	−0.0346	0.9153

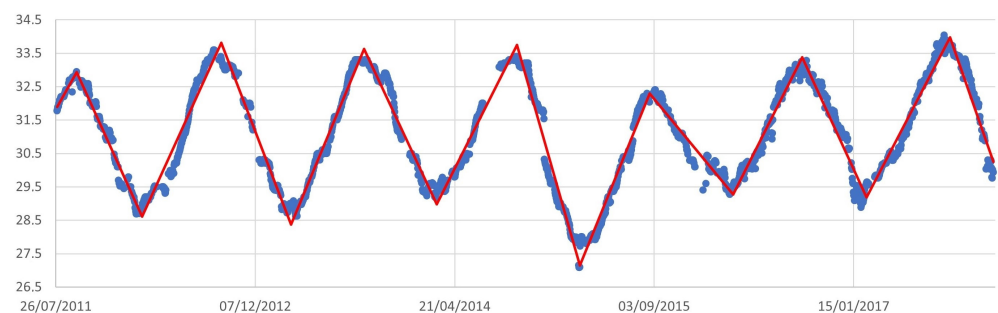


Figure 3. The scatter plot of the data series $X(t)$ and the linear approximation of the deterministic trend $r(t)$.

4. Analysis of the Stochastic Component

In this section, we focus on the investigation concerning the nature of the time series $B_H(t) = X(t) - r(t)$. To this aim, some preparatory interpolations were performed on the dataset, as is customary. Since the times of the initial series were not equally spaced, we collected the missing data by performing a linear interpolation covering the whole observation period. The prepared dataset obtained after this phase consisted of $N + 1 = 2348$ observations. A plot of $B_H(t)$ is shown in Figure 4.

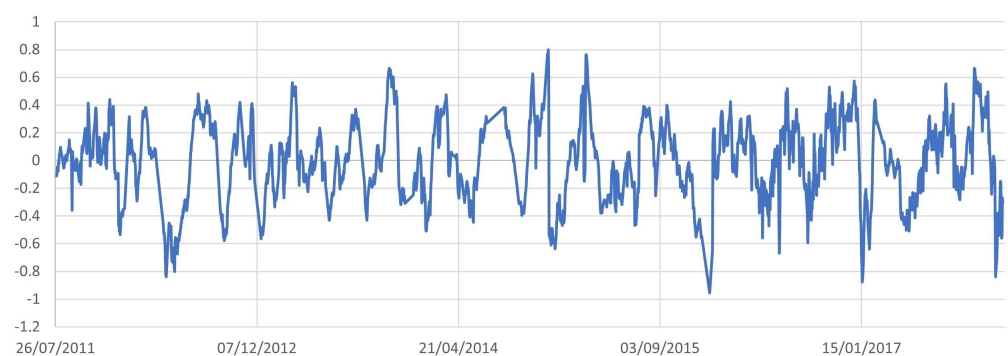


Figure 4. Plot of the time series $B_H(t) = X(t) - r(t)$.

As a preliminary study on the nature of the process $B_H(t)$, we represented the time series through a histogram and a qq-plot, as shown in Figure 5 and Figure 6, respectively. The obtained plots suggest classifying the stochastic component among Gaussian processes, thus justifying the following testing.

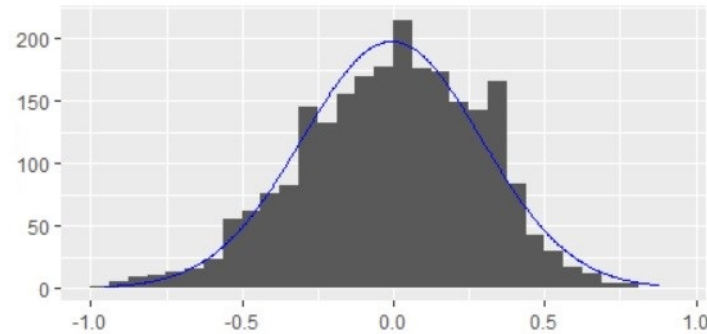


Figure 5. Histogram of the discrete time series $B_H(t)$.

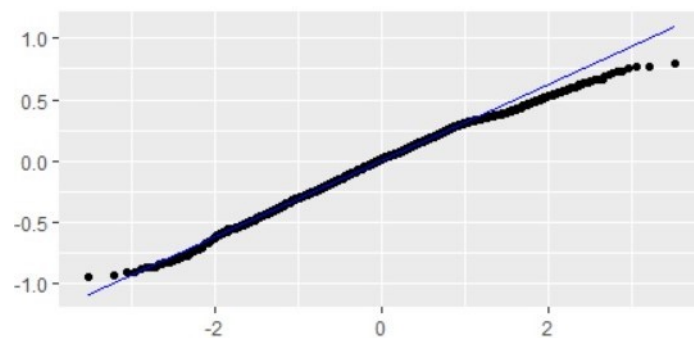


Figure 6. Qq-plot of the discrete time series $B_H(t)$.

4.1. Testing for Brownian Motion

With reference to the model in Equation (1), hereafter, we perform a test on the process $B_H(t)$, aiming to assess its suitable nature. With reference to the Brownian motion $B(t)$, we consider the following hypothesis:

Hypothesis 1. $B_H(t) = \sigma B(t)$.

Hypothesis 2. $B_H(t)$ is a confined or directed diffusion.

By adopting the test proposed by Briane et al. in [16], the asymptotic region of acceptance for $\mathcal{H}_0$ is given by

$$\left\{ q\left(\frac{\alpha}{2}\right) \leq \frac{S_D^N}{\hat{\sigma}\sqrt{t_N}} \leq q\left(1 - \frac{\alpha}{2}\right) \right\},$$

where

$$S_D^N = \max_{j=1, \dots, N} |B_H(t_j) - B_H(t_0)|,$$

$$\hat{\sigma} = \left\{ \frac{1}{N} \sum_{j=1}^N \frac{[B_H(t_j) - B_H(t_{j-1})]^2}{t_j - t_{j-1}} \right\}^{1/2},$$

and $q(\alpha)$ is the quantile of

$$\sup_{0 \leq s \leq 1} |B(s) - B(0)|.$$

The extremes of the acceptance interval for the significance level $\alpha = 0.05$ are given by

$$q(0.025) = 0.834, \quad q(0.975) = 2.940,$$

while the value obtained for the test statistic is 0.0787. The null hypothesis of the process $B_H(t)$ being Brownian motion is therefore rejected. This justifies the further analysis conducted hereafter.

Let us now conduct an investigation on $B_H(t)$ to verify whether it might follow an fBm or fGn trend. Consequently, we also estimate the Hurst exponent H of the process. To this aim, we apply the theoretical procedure consisting of evaluating the Fourier spectrum $S(f)$ of the data and studying its behavior with respect to the frequencies f . In fact, a relation analogous to Equation (3) for long memory processes can be presented in the frequency domain of the time series. In particular, for an fBm or a fGn process, the estimated Fourier spectrum $S(f)$ and the frequencies f are asymptotically evaluated as

$$S(f) = S(f_0) f^{-\beta}, \quad (4)$$

where both $S(f_0)$ and β are constants. The coefficient β is linked to the Hurst exponent H by different equations, depending on the nature of the underlying process. Indeed, for an fGn (fBm) process, Equation (4) holds with $\beta = 2H - 1$ ($\beta = 2H + 1$) (cf. [17–19]). Recalling that $0 < H < 1$, we find the following:

- (i) If $-1 < \beta < 1$, then the time series can be identified as a realization of fractional Gaussian noise;
- (ii) If $1 < \beta < 3$, then the data series represents a sample path of fractional Brownian motion.

From Equation (4), we have

$$\log S(f) = \log S(f_0) - \beta \log f,$$

so that an estimation of the coefficient β can be obtained by plotting on a log-log scale the frequencies and the values of $S(f)$ and then taking the opposite value of the slope of the least squares line as an estimation.

To attain an estimation of the function $S(f)$, we evaluate the periodogram of the data series, which for $B_H(t)$ is defined by

$$I(f) = \left| \frac{1}{2\pi N} \sum_{i=0}^N B_H(t_i) e^{ift_i} \right|^2.$$

In order to achieve the best possible approximation for the Fourier spectrum, many modified versions of the periodogram have been devised. In particular, we chose the estimation of $S(f)$ using the modified periodogram method suggested by Welch in [20]. The technique consists of dividing the signal into overlapping segments and averaging the modified periodograms calculated in each window to obtain the final result. The modified Welch's periodogram is shown in Figure 7.

Hence, by applying a log-log transformation on the frequencies and the estimated periodogram, we obtained the plot shown in Figure 8. The estimate of β obtained as the slope of the regression line is given by

$$\beta = 1.853,$$

with the corresponding $R^2 = 0.907$. Since the coefficient β belongs to the interval $(1, 3)$, we could conclude that the data series followed an fBm trend. This also led to an estimation of the Hurst parameter as follows:

$$\hat{H} = \frac{\beta - 1}{2} = 0.427. \quad (5)$$

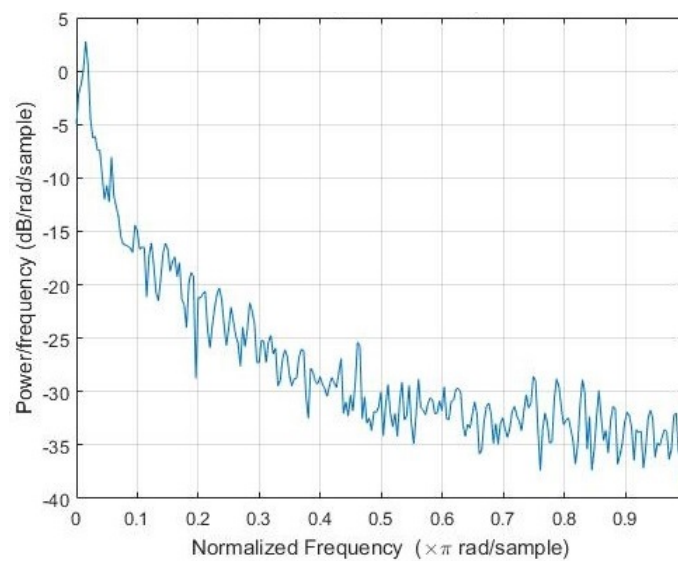


Figure 7. Modified Welch periodogram for the data series $B_H(t)$ plotted for different values of the frequency.

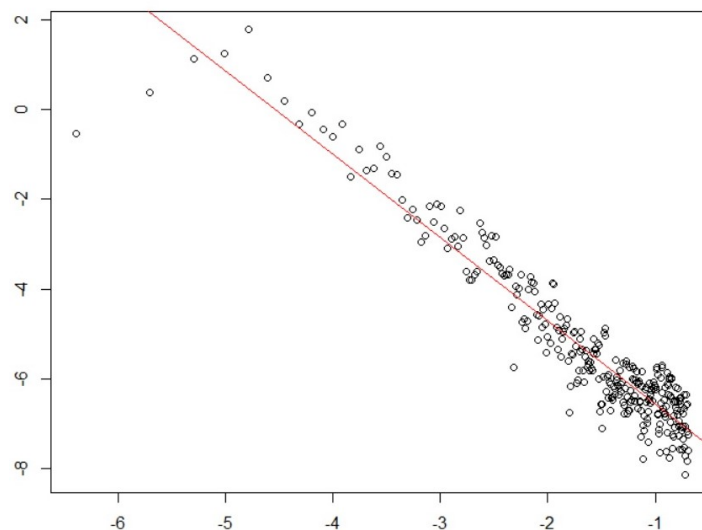


Figure 8. Log-log plot for the Welch's periodogram of the data series with respect to frequency. The slope of the least squares line estimates the coefficient β .

The estimation obtained in Equation (5) shows that the Hurst exponent of the process was less than 0.5, implying anti-persistence in the data trend. This result has already emerged in the literature in investigations related to soil temperatures. In the work by Zhang et al. [21], surveys of the soil temperature at different levels of depth showed that a weak anti-persistence ($H \sim 0.4$) was observed near the soil surface, consistent with our result.

In order to complete the estimation of the parameters for $B_H(t)$, we also evaluated the diffusion constant D of the process. This was performed directly from the data, calculating the sample variance of the increments $B_H(t) - B_H(t-1)$ of the data series. The estimate obtained to this point was

$$\hat{D} = 0.0846. \quad (6)$$

To further confirm the hypothesis suggested from the periodogram, we performed analysis of the residuals of the time series making use of simulations of fBm. To this aim, we denoted with $B_{\hat{H}, \hat{D}}(t)$ a simulated sample path of a fractional Brownian motion process with a Hurst exponent and scale parameter estimated as in Equations (5) and (6). Then, the residual

$$z(t) = X(t) - r(t) - B_{\hat{H}, \hat{D}}(t)$$

was evaluated by means of the residuals plot. Moreover, the normality of the residuals was assessed hereafter through two tests: the Shapiro–Wilk test and robust Jarque–Bera test for Gaussianity.

4.2. The Shapiro–Wilk Test

Since the test statistic W of the Shapiro–Wilk test tended to detect even small departures from the null hypothesis when the sample size was sufficiently large, it was convenient for reducing the sample to a smaller size. Hence, from now on, we consider a subset $\tilde{X}(t)$ of the main dataset which contains weekly spaced observations. After simulating a sample path $\tilde{B}_{\hat{H}, \hat{D}}$ of fBm of a length $M = \lfloor (N + 1)/7 \rfloor$, we plotted the residual series

$$\tilde{z}(t) = \tilde{X}(t) - \tilde{r}(t) - \tilde{B}_{\hat{H}, \hat{D}}(t)$$

with respect to the values of $\tilde{X}(t)$ (see Figure 9). Notice that the residuals do not seem to follow a specific path; rather, they appear to be distributed randomly with respect to $\tilde{X}(t)$.

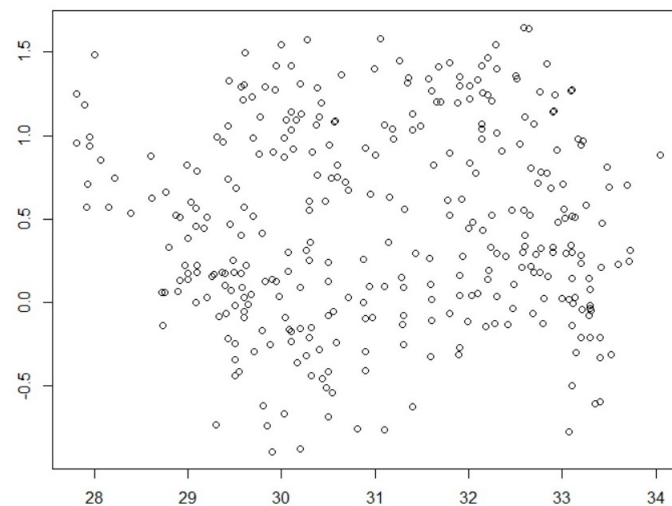


Figure 9. Plot of the residuals series $\tilde{z}(t)$ with respect to the temperatures $\tilde{X}(t)$, equally spaced with one observation for each 7 days.

Let us now describe the adopted testing strategy. Since we were interested in studying the behavior of the residual series, we first simulated the fBm path $\tilde{B}_{\hat{H}, \hat{D}}(t)$ for a total of $n = 10^4$ iterations. For each simulation, we evaluated the residuals series $\tilde{z}_i(t)$ and first performed a Shapiro–Wilk test for $i = 1, \dots, n$. The procedure was necessarily based on the repetition of the Shapiro–Wilk test since we dealt with simulated fBm paths. Hence, we considered the collection of test statistics

$$W_i = \frac{\sum_{j=1}^M a_j \tilde{z}_i(t_{(j)})}{\sum_{j=1}^M (\tilde{z}_i(t_j) - \mu_i)^2}, \quad i = 1, \dots, n, \quad (7)$$

where W_i refers to the i th simulation and where μ_i represents the sample mean of $\tilde{z}_i(t)$. The coefficients a_j , as is well known, are related to some moments of the order statistics of i.i.d. random variables sampled from the standard normal distribution. The values of W_i , as in Equation (7), ranged between 0 and 1. The i th test rejects the Gaussian hypothesis when W_i is close to 0, and hence we accepted the null hypothesis if $W_i \geq 0.98$ for $i = 1, \dots, n$. As shown in Figure 10, a large proportion of the values of W_i was greater than 0.98. This result was observed for more than 70% of the values of W_i even for various repetitions of the n -simulation procedure (cf. [22,23]). This was also confirmed by the cumulative

histogram of W_i shown in Figure 11. The results of the test suggest the acceptance of the Gaussianity hypothesis.

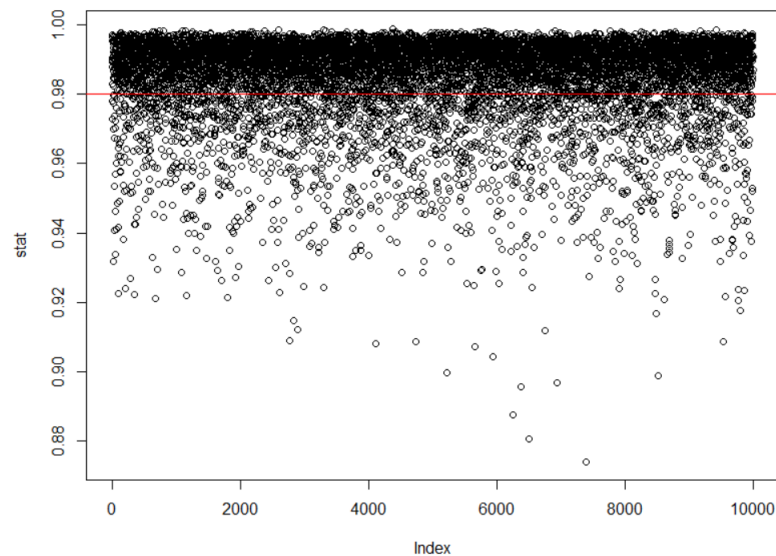


Figure 10. The values of the statistic W_i evaluated according to Equation (7).

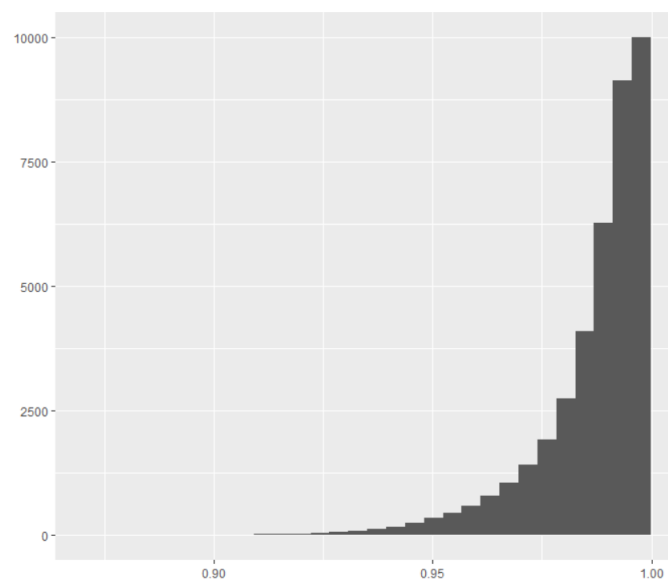


Figure 11. Cumulative histogram for the W_i statistic.

4.3. Robust Jarque–Bera Test

Let us now perform a modified version of the Jarque–Bera test. The statistic for the original Jarque–Bera test for a sample of a size n is given by (cf. [24])

$$JB = \frac{n}{6} \left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \right)^2 + \frac{n}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2} - 3 \right)^2,$$

which is a combination of the sample skewness $\left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \right)$ and the sample kurtosis $\left(\frac{\hat{\mu}_4}{\hat{\mu}_2} - 3 \right)$ obtained from the data through the estimates $\hat{\mu}_i$ of the central moments $i = 2, 3, 4$. If the data series is distributed normally, one has that $JB \sim \chi^2(2)$ asymptotically. Thus, under the null hypothesis of the time series being Gaussian, both the expected skewness and the

expected kurtosis are zero. Therefore, higher values of the JB statistic lead to the rejection of the Gaussianity hypothesis.

For our testing procedure, we followed the modified version of the test proposed by Gel and Gastwirth in [25], which is known as the robust Jarque–Bera (RJB) test. The modification consisted of substituting the estimate of the spread, formerly $\hat{\mu}_2$, with the average absolute deviation from the sample median, henceforth denoted as Med , given by

$$J_M = \frac{\sqrt{\pi/2}}{M} \sum_{j=1}^M |z(\tilde{t}_j) - Med|,$$

where M refers to the sample size. This substitution makes the statistics less influenced by outliers. Therefore, the statistic becomes

$$JB_M = \frac{M}{C_1} \left(\frac{\hat{\mu}_3}{J_M^3} \right)^2 + \frac{M}{C_2} \left(\frac{\hat{\mu}_4}{J_M^4} - 3 \right)^2,$$

where C_1 and C_2 are positive constants estimating moments of J_M , which is obtained through Monte Carlo simulations. Moreover, the p -values for the RJB test are obtained by use of $k = 10,000$ Monte Carlo simulations instead of the $\chi^2(2)$ distribution, thus obtaining a better approximation (cf. [26]).

To perform the RJB test, we repeated 10 blocks of n simulations with 3 different numbers of iterations ($n = 100, 500$ and 1000). For each simulation, the residual series

$$\{\tilde{z}_i(t_j), j = 1, \dots, M\}_{i=1, \dots, n} \quad (8)$$

was tested. The choice of performing blocks of simulations arose from the need for reducing the computational times. The results, reported in Table 4, showed compliance with the Shapiro–Wilk test and led to the acceptance of the Gaussianity hypothesis for the residuals series.

Table 4. Percentages of acceptable values for the residuals series in Equation (8) tested with the RJB method and repeated for $h = 1, \dots, 10$, with significance level $\alpha = 0.02$ and different values of n .

h	$n = 100$	$n = 500$	$n = 1000$
1	64%	66.8%	68.0%
2	76%	72.2%	69.9%
3	68%	70.0%	66.1%
4	66%	69.4%	71.3%
5	70%	71.0%	70.5%
6	68%	71.4%	71.1%
7	68%	70.8%	70.5%
8	71%	67.0%	69.9%
9	70%	73.0%	68.1%
10	70%	69.4%	71.5%

5. Statistical Test for the Model

To further confirm the results obtained in Section 4, we performed a statistical test on the whole N -term data series for the hypothesis of $B_H(t)$ being fractional Brownian motion with parameters $\hat{H} = 0.427$ and $\hat{D} = 0.0846$. To this aim, we used the inference test presented in Sikora [27] based on the detrending moving average (DMA) of the data. The hypotheses for the test were

$\mathcal{H}_0 : \{B_H(t_1), B_H(t_2), \dots, B_H(t_N)\}$ is a trajectory of fBm with parameters $\hat{H}$ and $\hat{D}$,

$\mathcal{H}_1 : \{B_H(t_1), B_H(t_2), \dots, B_H(t_N)\}$ is not a trajectory of fBm with parameters $\hat{H}$ and $\hat{D}$.

The corresponding test statistic for a fixed value $m > 1$ was given by

$$S^2(m) = \frac{1}{N-m} \sum_{j=m}^N (B_H(t_j) - \bar{B}_H^m(t_j))^2, \quad (9)$$

where $\bar{B}_H^m(t_j)$ denotes the moving average of the m observations $B_H(t_{j-i})$, $i = 0, 1, \dots, m-1$.

The p -value for this test was defined by an infinite series, so it needed to be computed as an empirical quantile obtained from series of generalized chi-squared random variables. Recalling Step 5 in Section 3 of [27], the p -value of the test was calculated as

$$p = \frac{2}{L} \min\{\#(\sigma_l^2(m) < S^2(m)), \#(\sigma_l^2(m) > S^2(m))\}. \quad (10)$$

In this formula, L is the number of samples generated for the chi-squared distribution. For a good estimate of p , we set $L = 1000$. Moreover, in Equation (10), we considered

$$\sigma_l^2(m) = \frac{1}{N-m} \sum_{j=1}^{N-m+1} \lambda_j(m) U_j^l, \quad l = 1, 2, \dots, L,$$

where $\mathbf{U}^l = \{U_1^l, U_2^l, \dots, U_{N-m+1}^l\}$ is the l th chi-square sample and $\lambda_j(m)$ is the j th eigenvalue of the sample matrix

$$\tilde{\Sigma} = \{\mathbb{E}[B_H(t_j)B_H(t_k)]; j, k = 1, 2, \dots, N-m+1\}.$$

Recalling Equation (9), since the search for an optimal value for m was still an open question, we divided the data series into 10 subsets of about 250 observations each in order to adapt the procedure even to cases with large numbers of observed data. Then, we ran the test in every subset for each suitable value of m .

After that, we collected the p -values and created a box plot with jitter for each subset in order to determine the outlier values for p (cf. Figure 12).

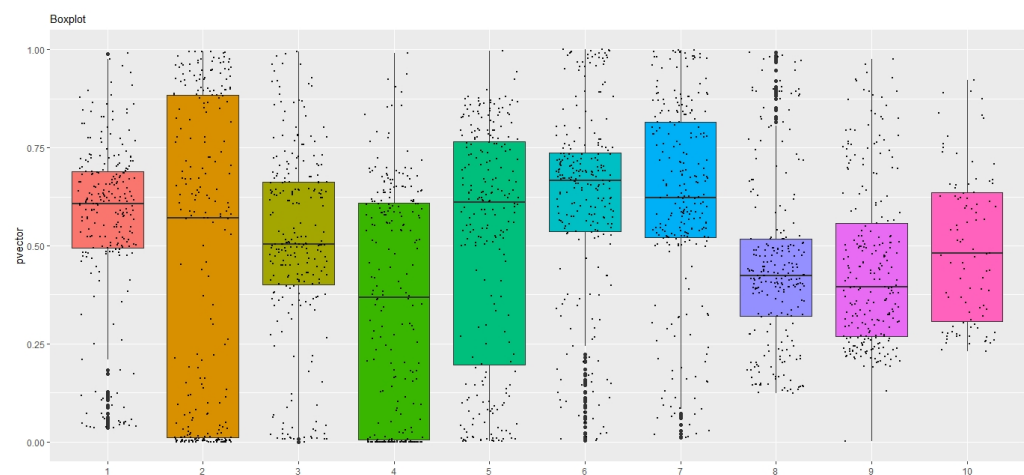


Figure 12. Box plot for the values of p obtained in the 10 subsets of the data.

The significance level considered for this test was $\alpha = 0.02$. Therefore, the results in each subset were collected and divided into three categories:

- (i) For $p \leq 0.02$, the null hypothesis is rejected;
- (ii) For $0.02 < p \leq 0.05$, the values are considered as a “warning”;
- (iii) For $p > 0.05$, the hypothesis $\mathcal{H}_0$ cannot be rejected.

Table 5 shows the results, with the percentages of the number of rejected, warning and acceptable values over the total p -values calculated in the 10 datasets. It can be observed

that the null hypothesis could not be rejected in almost 90% of the cases covered, considering all the subsets and all the different choices of m for the test statistic. This allowed stating the validity of the hypothesis that the time series $B_H(t)$ followed an fBm trend.

Table 5. Number of cases in which the DMA statistic test had different outcomes for the data series $B(t)$.

Set	Reject Values ($p \leq 0.02$)	Perc.	Warning Values ($0.02 \leq p \leq 0.05$)	Perc.	Acceptable Values ($p > 0.05$)	Perc.
1	0	0%	12	4.86%	235	95.14%
2	70	28.23%	12	4.84%	154	62.10%
3	17	6.85%	6	2.42%	219	88.31%
4	70	28.23%	6	2.42%	166	66.94%
5	23	9.27%	10	4.03%	205	82.66%
6	4	1.61%	4	1.61%	236	95.16%
7	3	1.21%	3	1.21%	239	96.37%
8	0	0%	0	0%	248	100%
9	1	0.4%	0	0%	247	99.6%
10	0	0%	0	0%	95	100%

6. Conclusions

The performed analysis shows that the proposed stochastic model is suitable to fit the observed temperatures. Indeed, the obtained results suggest that the coefficients obtained for the process $X(t)$ display high values for R^2 , thus ensuring the goodness of the fit. In conclusion, this study reveals that the evolution of the temperatures in the Campi Flegrei caldera can be modeled by the sum of a deterministic component, which represents the seasonal trend, and fractional Brownian motion.

Possible future developments of the performed analysis can be oriented toward (1) the prevision of the trend of the seasonal terms and (2) the study of other quantities of interest, such as leaked radon.

Author Contributions: Conceptualization, A.D.C., B.M. and V.M.; Investigation, A.D.C., B.M. and V.M.; Methodology, A.D.C., B.M. and V.M.; Writing, review and editing, A.D.C., B.M. and V.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Italian MIUR-PRIN 2017, project “Stochastic Models for Complex Systems”, No. 2017JFFHSH.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are members of the research group GNCS of INdAM (Istituto Nazionale di Alta Matematica).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

OLS	ordinary least squares
fBm	fractional Brownian motion
fGn	fractional Gaussian noise
RJB	robust Jarque–Bera test
DMA	detrending moving average
i.i.d.	independent and identically distributed
LRD	long-range dependency

References

- Sebastiani, G.; Malagnini, L. Forecasting the Next Parkfield Mainshock on the San Andreas Fault (California). *J. Ecol. Nat. Resour.* **2020**, *4*, 1–45. [\[CrossRef\]](#)
- Giordano, L.M.; Morale, D. A fractional Brownian–Hawkes model for the Italian electricity spot market: Estimation and forecasting. *J. Energy Mark.* **2021**, *14*, 1–44. [\[CrossRef\]](#)
- Sabbarese, C.; Ambrosino, F.; Chiodini, G.; Giudicepietro, F.; Macedonio, G.; Caliro, S.; De Cesare, W.; Bianco, F.; Pugliese, M.; Roca, V. Continuous radon monitoring during seven years of volcanic unrest at Campi Flegrei caldera (Italy). *Sci. Rep.* **2020**, *10*, 9551. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nourdin, I. *Selected Aspects of Fractional Brownian Motion*; Bocconi & Springer Series; Springer: Berlin/Heidelberg, Germany, 2012.
- Prakasa Rao, B.L.S. *Statistical Inference for Fractional Diffusion Processes*; Wiley: Hoboken, NJ, USA, 2010. [\[CrossRef\]](#)
- Mattia, F.; Le Toan, T.; Davidson, M.; Borgeaud, M. On the scattering from natural rough surfaces. *IGARSS* **1999**, *5*, 2413–2415. [\[CrossRef\]](#)
- Yin, Z.M.; Ranalli, G. Modelling of earthquake rupturing as a stochastic process and estimation of its distribution function from earthquake observations. *Geophys. J. Int.* **1995**, *123*, 838–848. [\[CrossRef\]](#)
- Taqqu, M.S.; Teverovsky, V.; Willinger, W. Estimators for longrange dependence: An empirical study. *Fractals* **1996**, *3*, 785–798. [\[CrossRef\]](#)
- Travaglino, F.; Di Crescenzo, A.; Martinucci, B.; Scarpa, R. A New Model of Campi Flegrei Inflation and Deflation Episodes based on Brownian Motion Driven by Telegraph Process. *Math. Geosci.* **2018**, *50*, 961–975. [\[CrossRef\]](#)
- Mandelbrot, B.B.; Van Ness, J.W. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Rev.* **1968**, *10*, 422–437. [\[CrossRef\]](#)
- Dieker, T. Simulation of Fractional Brownian Motion. Ph.D. Thesis, University of Twente, Enschede, The Netherland, 2004.
- Li, M. On the Long-Range Dependence of Fractional Brownian Motion. *Math. Probl. Eng.* **2013**, *2013*, 842197. [\[CrossRef\]](#)
- Balcererek, M.; Burnecki, K.; Thapa, S.; Wylomańska, A.; Chechkin, A. Fractional Brownian motion with random Hurst exponent: Accelerating diffusion and persistence transitions. *arXiv* **2022**, arXiv:2206.03818.
- Han, D.; Korabel, N.; Chen, R.; Johnston, M.; Gavrilova, A.; Allan, V.J.; Fedotov, S.; Waigh, T.A. Deciphering anomalous heterogeneous intracellular transport with neural networks. *eLife* **2019**, *9*, e52224. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hudson, D.J. Fitting Segmented Curves Whose Join Points Have to Be Estimated. *J. Am. Stat. Assoc.* **1966**, *61*, 1097–1129. [\[CrossRef\]](#)
- Briane, V.; Vimond, M.; Kervrann, C. An adaptive statistical test to detect non Brownian diffusion from particle trajectories. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016.
- Roume, C.; Ezzina, S.; Blain, H.; Delignieres, D. Biases in the Simulation and Analysis of Fractal Processes. *Comput. Math. Methods Med.* **2019**, *2019*, 4025305. [\[CrossRef\]](#) [\[PubMed\]](#)
- Montanari, A.; Taqqu, M.S.; Teverovsky, V. Estimating Long-Range Dependence in the Presence of Periodicity: An Empirical Study. *Math. Comput. Model.* **1999**, *29*, 217–228. [\[CrossRef\]](#)
- Flandrin, P. On the spectrum of fractional Brownian motions. *IEEE Trans. Inf. Theory* **1989**, *35*, 197–199. [\[CrossRef\]](#)
- Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [\[CrossRef\]](#)
- Zhang, T.; Shen, S.; Cheng, C.; Song, C.; Ye, S. Long-Range Correlation Analysis of Soil Temperature and Moisture on A'rou Hillsides, Babao River Basin. *J. Geophys. Res. Atmos.* **2018**, *123*, 12–606. [\[CrossRef\]](#)
- Royston, P. Approximating the Shapiro-Wilk W-test for non-normality. *Stat. Comput.* **1992**, *2*, 117–119. [\[CrossRef\]](#)
- Royston, P. Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *J. R. Stat. Soc.* **1995**, *44*, 547–551. [\[CrossRef\]](#)
- Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *Int. Stat. Rev.* **1987**, *55*, 163–172. [\[CrossRef\]](#)
- Gel, Y.R.; Gastwirth, J.L. A robust modification of the Jarque–Bera test of normality. *Econ. Lett.* **2008**, *99*, 30–32. [\[CrossRef\]](#)
- Deb, P.; Sefton, M. The distribution of a Lagrange multiplier test of normality. *Econ. Lett.* **1996**, *51*, 123–130. [\[CrossRef\]](#)
- Sikora, G. Statistical test for fractional Brownian motion based on detrending moving average algorithm. *Chaos Solitons Fractals* **2018**, *116*, 54–62. [\[CrossRef\]](#)